第29卷第2期
2009年3月

云南师范大学学报
Journal of Yunnan Normal University

Vol. 29 No. 2
Mar. 2009

基于MATLAB的多元非线性回归模型*

董大校
(临沧师范高等专科学校,云南 临沧 6770000)

摘　要： MATLAB是源于矩阵运算的一种高度集成的计算机语言。它提供了强大的科学运算、灵活的程序设计流程、高质量的图形可视化与界面设计、便捷的与其他程序和语言接口的功能。文章充分利用MATLAB统计工具箱的优势,通过程序的实现,对多元非线性回归模型的未知参数的估计方法以及对估计后的模型预报做出研究,并以实例验证了该方法的有效性。

关键词： MATLAB;多元非线性回归;最小二乘法;统计工具箱

中图分类号:TP301　　**文献标识码：** A　　**文章编号：** 1007-9793(2009)02-0045-04

1　预备知识

非线性回归最小二乘法拟合的基本原理[1]。

对给定数据 $(x_i,y_i)(i=0,1,\cdots,m)$,在取定的函数类 Φ 中,求 $p(x)\in\Phi$,使误差 $r_i=p(x)-y_i(i=0,1,\cdots,m)$ 的平方和最小,即

$$\sum_{i=1}^{m} r_i^2=\sum_{i=0}^{m}[p(x_i)-y_i]^2$$

最小,从几何意义上讲,就是寻求与给定点 $(x_i,y_i)(i=0,1,\cdots,m)$ 的距离平方和为最小的曲线 $y=p(x)$(图1)。函数 $p(x)$ 称为拟合 函数或最小二乘解,求拟合函数 $p(x)$ 的方法称为曲线拟合的最小二乘法。

2　MATLAB非线性曲线拟合命令介绍

2.1 nlinfit函数[2]

用nlinfit函数进行非线性最小二乘数据拟合。该函数使用高斯—牛顿算法,调用格式如下:

●beta = nlinfit(X,y,fun,beta0)　用最小二乘法估计非线性函数系数。y 为响应值(因变量)矢量。一般地,为自变量值组成的设计矩阵,每一行对应与 y 中的一发个值。但是,X 可以是fun参数能接受的任何数组。fun参数为一函数,该函数具有下面的形式

yhat = myfun(beta,X)

其中beta为系数矢量,X为设计矩阵。fun为参数返回一个拟合y值的yhat矢量。beta0为包含系数初始值的矢量。

●[beta,r,J] = nlinfit(X,y,fun,beta0)　返回拟合系数(beta)、残差(r)、和雅可比矩阵J,这些参数可以用于nlintool函数,生成预测值的误差估计;或用于nlparci函数生成系数的误差估计。

2.2 nlintool函数

使用nlintool函数可以对数据进行非线性方程拟合并交互图形显示,其调用格式如下:

---

* 收稿日期:2008-10-23

作者简介:董大校(1965-),男,云南省临沧市人,副教授,主要从事数学教学科研工作.

●nlinfit(x,y,fun,beta0)　为一预测图,它提供数据(x,y)的非线性曲线拟合。它用两条红色曲线来表示预测值的95%置信区间。beta0为一矢量,包含参数的初值。fun参数为一函数,该函数具有yhat = myfun(beta,X)的形式:

其中beta为系数矢量,X为设计矩阵。

nlintool函数显示一个图形“矢量”,其中的每个图形对应于输入矩阵x的每一列。响应变量y为一列矢量,与x的行数相匹配。

●nlintool(x,y,FUN,beta0,alpha)　用图形表示预测值的100(1 - alpha)%置信区间。当x为矩阵时,将为每一列单独生成图形,应变量y为一列变量,对应于x的行。alpha()的默认值为0.05生成95%置信区间。

2.3 nlpredci 函数

可利用nlpredci函数计算非线性模型预测值的置信区间,其调用格式如下:

●ypred = nlpredci(FUN,inputs,beta,r,J)　给定拟合参数(beta)、残差 (r)和和雅可比矩阵(J),返回预测响应。输入是非线性函数中独立变量的数值矩阵。

2.4 nlparci 函数

可利用nlparci函数计算非线性模型中参数估计值的置信区间。

●nlparci(beta,r,J)　给定拟合参数(beta)、残差(r)、和解处的雅可比矩阵(J),返回非线性最小二乘参数估计(beta)的95%置信区间ci。nlparci函数使用nlinfit函数的输出作为输入。

2.5 regress 函数

用regress函数进行多元线性回归。

●[b,bint,r,rint,stats] = regress(y,x,alpha)　给出bint和rint的100(1 - alpha)%置信区间。

2.6 stepwise 函数

用stepwise函数进行逐步回归,它使用交互环境进行分析。其调用格式如下:

●stepwise(x,y,inmodel,penter,premove)　指定模型的初始状态和要使用的置信区间。inmodel为长度为x中列数的逻辑矢量。inmodel指定包含在初始模型中的自变量。默认时不包含x中的列。penter指定引入自变量的最大p值,默认时为0.05。premove指定剔除自变量的最小p值,默认时为0.10。

3 实例验证

3.1 反应动力学模型

反应动力学中的Hougen - Watson模型是非线性模型的一个典型例子。其模型如下:

$$rate = \frac{\beta_1 \cdot x_2 - x_3/\beta_5}{1 + \beta_2 \cdot x_1 + \beta_3 \cdot x_2 + \beta_4 \cdot x_3}$$

其中:$\beta_1, \beta_2, \cdots, \beta_5$为未知参数,$x_1, x_2$和$x_3$为三个输入变量。三个输入为氢(hydrogen)、n - 戊烷(n - pentane)、异戊烷(isopentane)。上述模型显然为多元非线性模型。文件reaction.mat包含了反应的仿真数据。

通过计算可知,反应动力学Hougen - Watson模型中的5个未知参数为

$$\beta_1 = 1.2526, \beta_2 = 0.0628, \beta_3 = 0.04, \beta_4 = 0.1124, \beta_5 = 1.1914$$

最终模型为:$rate = \dfrac{1.2526 \cdot x_2 - x_3/1.1914}{1 + 0.0628 \cdot x_1 + 0.04 \cdot x_2 + 0.1124 \cdot x_3}$

3.2 财政收入预测[3]

财政收入与国民收入、工农业总产值、人口、就业人口、固定资产投资等因素有关。采用逐步回归分析构造预测模型。以财政收入作为因变量y,自变量为国民收入(x_1)、工业总产值(x_2)、农业总产值

(x_3)、总人口(x_4)、就业人口(x_5)、固定资产投资(x_6)。其样本数 $n = 30$,自变量数 $p = 6$ 。

首先编写非线性函数的M函数文件model. m代码。然后在命令窗口输入由 x_1、x_2、x_3、x_4、x_5 样本数据构成的矩阵 x 和财政收入样本数据 y , 取 beta0 = [0.50 -0.03 -0.60 0.01 -0.02 0.35];用交互式非线性拟合图形工具来显示,在MATLAB命令窗口中输入如下命令:

nlintool(X,Y,'model',beta0,0.01);得到 nlintool 分析窗口。

由此得到财政收入与各因素之间关系的回归模型1为:

$$y = 0.5245x_1 - 0.0293x_2 - 0.631x_3 + 0.0113x_4 - 0.0231x_5 + 0.3652x_6 \qquad (1)$$

对该模型进行多元线性回归检验,输入命令:> > [b,bint,r,rint,stats] = regress(y,x)

stats =0.9805　301.0305　　0　951.5530

结果分析:bint为各系数的置信区间。stats矢量是值分别为相关系数的平方、F 值和显著性概率 p。相关系数平方值 $R^2 = 0.9805$,说明模型拟合程度较高。显著性概率 $p = 0$,小于0.05,故拒绝零假设,认为模型1中至少有一个自变量的系数不为零,因而从总体上模型1是有意义日的。

结果 b 给出了回归系数的估计值,检查它们的置信区间发现,x_2 系数的置信区间包含了零点,表明回归变量 x_2 对变量 Y 的影响不是太显著,因此从模型中移出次变量,对该模型进行逐步回归检验,在MATLAB命令窗口输入并运行命令:> > stepwise(x,y)

除了 x_1,x_3,x_6 外,其他自变量的回归系数置信区间都包含零点,保留 x_1,x_3,x_6 ,将其他变量移出回归模型,R—square =0.979645,是指因变量的97.96%可由模型确定,$F = 417.104$ 远远超过了 F 检验的临界值,$p = 0$ 远小于显著性水平,这些说明用包含 x_1,x_3,x_6 的线性模型来描述 y 是合适的。x_1 的回归系数为0.382869,x_3 的回归系数为 -0.496975,x_6 的回归系数为0.408624,常数项由Intercep给出,即 $b = 177.07$,利用逐步回归得到的回归模型2为:

$$y = 0.382869x_1 - 0.496875x_3 + 0.408624x_6 + 177.07 \qquad (2)$$

由于 x_1,x_3,x_6 对 y 具有较明显的显著性,用更为精确的拟合曲线来拟合表示 x_1,x_3,x_6 和 y 的关系。做 x_1,x_3,x_6 分别对 y 的散点图,具体步骤在MATLAB命令窗口输入命令:

> > x1 =x(:,1); x3 =x(:,3); x6 =x(:,6); > > subplot(1,3,1); plot(x1,y,'+'); > > subplot(1,3,2); plot(x3,y,'+'); > > subplot(1,3,3); plot(x6,y,'+')

再做经变换后的散点图,对比其线性效果较好的,得出如下拟合曲线:

$$y = -71.391 + 0.473x_1; y = 1620.46 - 6.509x_3 + 0.01x_3^2; y = 141.921 + 1.158x_6;$$

在MATLAB命令窗口输入如下命令:

```
> > subplot(1,3,1); > > plot(x1, -71.391 +0.473 * x1,'+');
> > subplot(1,3,2); > > plot(x3,1620.46 -6.509 * x3 +0.01 * x3.^2,'+')
> > subplot(1,3,3); > > plot(x6,141.921 +1.158 * x6,'+')
```

对应的 *R - Square* 分别为0.951,0.887,0.944。从相关系数的值可以看出,对应的拟合方程的拟合程度较高。所以,用拟合后的曲线来代替原来的 x_1,x_3,x_6 ,得到如下回归模型3:

$$y = 0.1811x_1 + 3.2342x_3 - 0.005x_3^2 + 0.4732x_6 - 597.4372 \qquad (3)$$

作为对比,在引入交互项,在MATLAB命令窗口输入下面命令:

> >x1 =x(:,1);x2 =x(:,2); x3 =x(:,3);x7 =x1. * x2; x8 =x1. * x3; > > x =[x x7 x8]; > > stepwise(x,y)

加入交互项后,利用逐步回归分析,从运算结果可得 *R - Square* = 0.98665,*F* = 461.923,*RMSE* = 27.9283,*p* =0.0000。各项参数值较模型(2)、(3)更好,所以,建立模型(4)

$$y = 0.845361x_1 - 0.0120252x_5 + 0.0000045x_7 - 0.00057x_8$$

即:

$$y = 0.845361x_1 - 0.0120252x_5 + 0.0000045x_1x_2 - 0.00057x_1x_3 \qquad (4)$$

从各统计量的数值看,模型(2)比模型(1)差一点,但模型(2)比模型(1)更为简单,模型(3)较模型

(1)、(2)虽然是非线性的回归模型,但实际上它的拟合效果并不好。模型(4)是在加入交互项后建立的,其系数中有的较小,几乎为零,代入数据进行预测,其效果不是很好。综合上述情形,建立的线性回归模型(2)更能真实的反映各自变量与因变量之间的关系。

4 结 语

本文从两方面入手,介绍了基于 MATLAB 的多元非线性回归模型参数的求解方法,同时也尝试了通过实验建立多元非线性回归模型的方法,虽然对于 3.2 财政收入预测模型并没有表现的像多元线性回归模型那样优秀,但是可以提供一种思想与方法,在以后的学习与实践中来完善这种思想与方法。

参 考 文 献:

[1] 袁建问. 经济计量学实验[M]. 北京:科学出版社,2002. 87 - 89.
[2] 苏金明,阮沈勇,王永利. MATLAB 工程数学[M]. 北京:电子工业出版社,2005. 50 - 83.

The Multivariate Nonlinear Regression Model Based on MATLAB

DONG Da - xiao
(Lin Cang Teachers' College)

Abstract: MATLAB is the origin of the matrix operations in a highly integrated computer language. It provided a powerful scientific computing, flexible program design process, high - quality graphics and visual interface design and convenient procedures and language and other interface features. In this paper, we take full advantage of MATLAB statistic tool box to achieve the multivariate nonlinear regression model. Estimate of the unknown parameters. examples are used to validate the method is effective.
Key words: MATLAB; Multivariate Nonlinear regression; Least squares; statistic tool box

(上接第 44 页)

Entire Solutions of Functional Equation $f^6(z)+g^6(z)+h^6(z)=1$

SU Min, LI Yu - hua
(Department of Mathematics, Yunnan Normal University, Yunnan Kunming 650092)

Abstract: This paper studies the entrie solutions of the functional equation $f^6(z)+g^6(z)+h^6(z)=1$. It is shown that there is no nonconstant entrie functions with order smaller than one satisfying the equation $f(z)$, $g(z)$, $h(z)$ with order smaller than one satisfying the equation $f^6(z)+g^6(z)+h^6(z)=1$. Meanwhile a new brief proof is given for the known results that if $n\geq 7$ there is no nonconstant entrie functions solutions of the functional equation $f^n(z)+g^n(z)+h^n(z)=1$.
Key words: Functional equation; Entire function; Value distribution theory

基于MATLAB的多元非线性回归模型



作者： 董大校， DONG Da-xiao
作者单位： 临沧师范高等专科学校,云南,临沧,6770000
刊名： 云南师范大学学报（自然科学版） ISTIC
英文刊名： JOURNAL OF YUNAN NORMAL UNIVERSITY(NATURAL SCIENCES EDITION)
年，卷(期)： 2009，29(2)
被引用次数： 0次

参考文献(2条)

1. 袁建问 经济计量学实验 2002

2. 苏金明. 阮沈勇. 王永利 MATLAB工程数学 2005

相似文献(7条)

1. 学位论文 康军 基于径向基函数神经网络的应用研究 2009

人工神经网络（ArtificialNeuralNetworks，简称为ANN）是理论化的人脑神经网络的模型，是模仿大脑神经网络结构和功能而建立起来的一种信息处理系统。人工神经网络的研究始于20世纪40年代，近二十年发展极为迅速，它的应用范围涉及到数学、工程、计算机、物理、生物、经济、管理等科学领域。本论文主要基于ANN的非线性逼近性质，借助MATLAB的神经网络工具箱，我们得出了如下的结果：

首先，我们实现了BP网络（下文中简写为BPNN）和径向基函数（RaclialBasisFunction）神经网络（下文中简写为RBFNN）的建立、学习和训练。通过仿真实例，体现了ANN的非线性逼近能力，并对BPNN和RBFNN进行比较研究，结果显示了RBFNN具有结构简单、训练速度快，并且有比较强的抗干扰能力，能达到很好的逼近效果。体现了RBFNN的函数逼近能力在多方面都优于BPNN。其次，我们把RBFNN作为一种新的回归方法应用到多元线性回归模型和多元非线性回归模型中。通过实例分析，结果表明RBFNN方法用于回归分析的拟合效果和预测效果很好而且简单方便。最后，我们用RBFNN进行了股票预测，通过以中国上证指数的540个交易日的数据为实验对象进行预测，得到了较好的预测效果。

本论文的结构安排如下：

第1章绪论。主要是论文的研究背景以及论文的主要工作。

第2章AAN基本理论。主要是ANN概述、BPNN和RBFNN简介以及ANN用于函数逼近的一些定理。为ANN的函数逼近能力的研究提供了理论基础。

第3章BPNN和RBFNN函数逼近仿真。主要是利用MATLAB神经网络工具箱，把BPNN和RBFNN应用于函数逼近的设计及仿真。

第4章RBFNN模型与多元回归模型。主要是RBFNN方法解决多元线性回归和多元非线性回归问题，并对传统回归方法和RBFNN方法在实际应用中进行了比较研究。

第5章RBFNN模型对股票市场价格进行预测。ANN可以看成是一个具有自学习功能的“黑箱”，在处理无法精确建模的问题上具有很大优势。由于股票的价格是非线性的时间序列，本文提出了基于RBFNN的股票价格预测模型，并应用该模型对于中国上证指数进行了预测。

结语。包括了论文的主要结论与不足，并为未来的研究提出建议。其中第3、4、5章是研究的核心部分。

2. 期刊论文 汪岚. 黄彩虹. Wang Lan. Huang Caihong 基于MATLAB色差预测多元回归模型的研究 -计算机与应用化学 2008,25(8)

由于间歇染色无法实现色差在线测量,直接影响染色的质量.在染色过程中,色差受多种工艺参数的影响,其中主要的有染料的初始浓度、染色时间、染色温度以及助剂的浓度.为了预测色差值,本文在实验数据分析的基础上,利用MATLAB统计工具箱的"nlinfit"函数,建立色差的多无非线性回归模型,并通过多元检验确定该模型的正确性.实验结果表明,多元非线性模型计算得到的色差值与实际值很接近,对生产实践具有一定的指导意义.

3. 学位论文 李宗波 船舶转向过程中速度下降问题的研究 2008

本文提出的船舶操纵性指数K'、T'估算公式较原有的线性估算公式大幅提高了K'、T'的预报精度，在保证回归方程显著性的基础上兼顾了数据拟合的精度和回归多项式的简洁性，符合精简实用的原则，对操船者在航海实践中评估船舶的操纵性能有一定的参考价值。

随着大型船舶的数量不断增长，船舶回转过程的速降问题已引起了人们的重视。原则上讲，回转中的速降需用非线性解析方法来确定，Abkowitz非线性数学模型以及分离型数学模型(MMG模型)虽然能预报包括速降在内的各种回转参数的变化情况，但都存在计算量大、较为复杂的问题。因此，实践中通常采用近似方法来估算之。戴维逊、费加耶夫斯基、Schoenherr等国外学者绘制了以相对回转直径为查表引数的速降系数曲线图或者给出了速降估算公式，但这些图表无法对船舶在不同载况(满载、半载、压载)下表现出的回转速降差异给出合理的解释。此外，这些图表和公式只能用于估算定常回转阶段速度V与初始速度V0之比，而对于船舶进入定常回转阶段之前的过渡阶段的速度下降情况无法进行估算，而过渡阶段恰好是船舶航行过程中进行转向避让经常涉及的范围，因而以上学者的研究成果存在一定的局限性。

本文另辟蹊径，从由响应型船舶运动数学模型推导出来的船舶操纵性指数K、T入手，用多项式回归的办法对47艘船舶的K、r指数统计资料进行分析，得出了具有一定精度的K'、T'指数估算公式。然后，以此为基础对多艘船舶的回转试验速降资料进行回归分析，把船舶的速降系数表征为船舶旋回性指数K'、船舶施加的舵角以及航向改变量△的函数关系，初步克服了国外关于回转速降研究成果的局限性，实现了对转向全过程的速降情况的估算。

本文以超大型油轮OPALIA号为例进行了回转速降的预报，用MATLAB绘制了速降曲线，并与国外学者的估算公式和曲线图进行了对比，取得了理想的效果，预报误差最大为14.6%，最小为0.3%。这表明，用多元非线性回归的方法实现对船舶整个转向过程中速降情况的预报具有一定的可行性。

4. 期刊论文 刘俏. 高伟星. 权春善. 范圣第. Liu Qiao. Gao Weixing. Quan Chunshan. Fan Shengdi 基于MATLAB发酵培养基的优化 -计算机与应用化学2006,23(7)

发酵培养基配比的优化是含有多变量非线性目标函数的优化.以青霉素发酵为例,通过考察青霉素效价产率的大小,用MATLAB统计工具箱的"regress"函数对培养基配比进行多元非线性回归,以含有约束的回归方程作为目标函数,调用MATLAB优化工具箱的"fmincon"函数,对其进行优化,确定了适宜的培养基配比.将优化条件得到的预测值与实测值比较,当培养基配比为乳糖8.0%、花生饼粉4.0%及苯乙酰胺0.213%时,目标产物的效价产率为1226.0,与优化结果1210.9比较接近.

5. 学位论文 刘涛 水库水质参数预测和富营养化预警 2005

当今全球水资源短缺已经成为不争的事实。水质富营养化问题作为全球性的水污染问题之一，正日益引起全社会的关注。尤其是作为城市饮用水水源的水库富营养化问题更是严重威胁着城市供水的安全。本文以于桥水库为研究对象，以保障城市安全供水为目标，系统地研究了水库富营养化问题。

首先，文章对于桥水库的水质现状进行了调查研究、收集资料等工作。并在此基础之上，对所收集的资料进行了筛选和分析，绘制出于桥水库水质参数的时空分布规律曲线。

其次，藻类恶性增殖作为富营养化的核心体现，同时也直接威胁着城市供水安全，所以本文以藻类为切入点，选择反映藻类数量的叶绿素浓度作为水质富营养化指标和预测预警对象进行研究，分别运用多元非线性回归方法、人工神经网络方法建立了预测模型，分别采用SAS和MATLAB编写了程序进行

模型的标定和验证求解。进而采用CE-QUALW2理论模型的方案加以辅助验证。并分别比较了三种方法的优劣。

最后，针对上述研究结果，采用当前流行的水质预警理论应用于于桥水库的实际情况，研究制定出了相应的分级预警方案。并结合2003年全年的预测数据给出预警级别，效果良好。

至此，本文结合于桥水库的实际情况系统地研究了水库富营养化问题，提出了一套完整的解决方案，实际分析的结果显示该方案切实可行，可以应用于实际。

6.期刊论文 李晓燕.张曙.余灯广.Li Xiaoyan.Zhang Shu.Yu Dengguang 三维打印成形粉末配方的优化设计 -机械科学与技术2006,25(11)

利用均匀设计方法对三维打印成形的粉末配方进行了试验研究,对试验数据进行回归分析,建立了粉末成分配方对制件性能的多元非线性回归方程.设计了粉末配方的多目标优化模型,利用MATLAB软件求解出最佳配方并通过试验加以验证.采用均匀设计试验进行回归分析和优化是安排多水平配方设计的良好方法.

7.学位论文 李琼 水库水质富营养化预测预警研究 2002

水质富营养化是全球性的水污染问题,正日益引起全社会的关注,尤其是作为城市饮用水水源的水库富营养化问题更是严重威胁着城市供水的安全.该文以于桥水库为研究对象,以保障城市安全供水为目标,系统地研究了水库富营养化问题.首先针对水库水质管理不完善的问题,引入了先进的GIS技术,研制开发了相关的水库水质管理软件,并结合于桥水库的具体地形图和水质信息建立了于桥水库水质管理信息系统,实现了水库水质的信息化管理,并为富营养化的预测、预警和防治规划提供了信息基础.其次,藻类恶性增殖作为富营养化的核心体现,同时也直接威胁着城市供水安全,所以该文以藻类为切入点,选择反映藻类数量的叶绿素浓度作为水质富营养化指标和预测预警对象进行研究,先后运用多元非线性回归方法和人工神经网络方法建立了预测模型,分别采用SAS和MATLAB编写了程序进行模型的标定和验证求解,并根据预测的结果结合水库的实际情况制定了相应的分级预警方案,在预警的基础上研究相应的应急防治措施,这些措施包括水库内的藻类削减措施和水厂采取的工艺和药剂方面的应急准备.最后,为了真正实现安全供水,对于桥水库富营养化成因进行了分析,并对现阶段可以实现的防治措施进行了系统的规划,使用Lingo软件对防治规划进行了优化分析,在保证水质、满足环境容量约束的条件下得到了最优的富营养化防治规划.